

# Data Explained

---

## **Annual Survey of Hours and Earnings linked to 2011 Census – England and Wales**

Exploring the nexus between immigration, integration and labour market outcomes

Author: Dr Ezgi Kaya

Date: October 2024

---

This Data Explained summarises experiences and learning from working with the Annual Survey of Hours and Earnings (ASHE) linked to 2011 Census – England and Wales dataset, while producing research exploring the links between immigration, integration, and labour market outcomes. This publication is intended to help guide future researchers using this data and to provide feedback into future dataset development and documentation.

The administrative data discussed in this Data Explained was made securely available through the ADR UK-funded Wage and Employment Dynamics programme. The data used in this research project comes from the Office for National Statistics (ONS) and was accessed through the ONS Secure Research Service (SRS). The data was not originally collected for research and it is expected that there are gaps and inconsistencies in its recording, a number of which are detailed in the following.

---

## Project details

[The project](#) aimed to provide new, high-quality evidence on labour market performance and integration of immigrant workers in England and Wales using the ASHE – 2011 Census linked dataset (ONS, 2023). With the immigrant population in England and Wales increasing from 7.5 million in 2011 to ten million in 2021 (ONS, 2022), understanding immigrant employees' experiences in the labour market is increasingly important. [Previous research](#) for the UK in this area has primarily relied on survey data, which depends on self-reported information, has limited details on employers, and often does not track individuals over time. The ASHE – 2011 Census linked dataset offers rich information to overcome these challenges.

By utilising this novel dataset, we compared labour market outcomes of immigrant (non-UK-born) and UK-born employees—including pay, hours worked, occupational skill level, and type of employment contract—and explored the main drivers of any differences. The large sample size allowed us to determine whether the findings vary based on other characteristics such as gender, ethnicity and [country-of-birth groupings](#). The linked employer-employee nature of the data facilitated the exploration of the role individual employers play in influencing labour market differences, through comparisons among observationally equivalent co-workers. Additionally, we investigated how factors related to immigration and integration such as English language proficiency, UK passport holding, and British national identity influence immigrant employees' performance.<sup>1</sup> The longitudinal element of the data further enabled a detailed examination of how labour market differences evolve over time by tracking the same de-identified individuals.

## Initial research questions

This project aimed to explore the following research questions:

1. How do labour market outcomes, including pay, hours of work, occupational skill level and type of employment contract, differ between UK-born and immigrant (non-UK-born) employees?
2. What are the main drivers of any differences in labour market outcomes; and what is the influence of other personal and work-related characteristics, or the role played by individual employers?
3. Does the labour market performance of immigrants vary by factors related to immigration and integration (English language proficiency, years of residence in the UK, year of arrival, holding a UK passport or British national identity)?
4. How do labour market outcomes of immigrant employees progress over time; and how do the labour market dynamics of immigrants compare to those born in the UK?

---

<sup>1</sup> Integration in this context is understood as mutual recognition between immigrants and the host population, although these analyses have focused only on areas where data is available. For a discussion of the measurement of integration, see Kierans (2021).

## Research methodology

The analysis used established econometric methods with four core elements:

- First, using cross-sectional data for 2011, we estimated regression models to compare the labour market outcomes of immigrants with that of UK-born employees. This was followed by decomposition analysis to differentiate the impact of observable characteristics from unobserved influences, including potential unequal treatment.<sup>2</sup>
- Second, utilising the linked employer-employee nature of the data, we examined how employers influence differences in labour market outcomes between immigrant and UK-born employees. We used regression analysis to estimate overall and within-employer gaps and applied decomposition methods to quantify the precise role employers play on the observed differences in the labour market outcomes considered.<sup>3</sup>
- Third, to investigate the role of factors related to immigration and integration, we estimated overall and within-employer gaps for different groups of immigrants (e.g. English proficient or not). This way, we identified whether the differences between UK-born and immigrant employees in labour market outcomes vary by these factors. Since these variables are only applicable to immigrants, in a separate exercise, we restricted the analysis to non-UK-born employees and made comparisons among them to further understand the role of these factors in determining their labour market performance.
- Finally, we utilised the longitudinal element of the data to track the same de-identified individuals from 2011 to 2018 and used regression analysis to examine how the labour market performance of immigrants progresses over time (i.e. since their year of arrival) compared to UK-born employees.

---

<sup>2</sup> See Kaya (2024a) for findings from these analyses.

<sup>3</sup> Findings from these analyses can be found in Kaya (2024b).

## Key variables

ASHE - 2011 Census linked dataset		
Variable name	Type	Description
aggcobpuk113_cen	String	Country of birth UK (indicator)
cob_cen	String	Country of birth
yrarr_year_cen	String	Year of arrival in UK
lrespuk11_cen	String	Length of arrival in UK
gpay_ashe	Numeric	Average gross weekly earnings for the reference period
thrs_ashe	Numeric	Average total paid hours worked during the reference period
occ10_ashe	Numeric	Occupation <a href="#">SOC 2010</a>
pt_ashe	Numeric	Permanent / temporary marker (type of employment contract)
entref_ashe	Numeric	Anonymised enterprise unit reference
mainlangprf11_cen	String	Proficiency in English
psspuk111_cen	String	Passports held (indicator for UK passport)
natidbpuk11_cen	String	British national identity
adr_id_ashe	Numeric	Anonymised personal identifier

Notes: The key variables included in the analysis are based on the 'Drop 2' of the ASHE - 2011 Census linked dataset constructed using the code files provided by the [Wage and Employment Dynamics](#) Team within the ONS SRS (see below). A discussion of the other variables included in the analysis can be found in Kaya (2024a).

## Summary of comments on specific variables

Variable names	Comments
aggcobpuk113_cen	Note that there is no single, universally accepted definition of a 'migrant', and different datasets and analyses use <a href="#">various definitions</a> . The ASHE - 2011 Census linked dataset does not include a specific migrant identifier. This project employed the definition based on the variable capturing whether country of birth is the UK, acknowledging that those in the non-UK-born category may hold British citizenship.
cob_cen	Although detailed information on the country of birth is captured by this variable, sample sizes for each country are often too small for separate analysis. Additionally, a tiny number of non-UK-born individuals (less than 1%) have country of birth information that is not specific enough (e.g. listed as 'Europe (Not otherwise specified)'). Researchers interested in country-specific analysis might consider using country-of-birth groupings constructed based on this variable.
lrespuk11_cen, yrarr_year_cen	The length of residence information pertains only to usual residents who were not born in the UK and is calculated based on the date of their last arrival to live in the UK, excluding short visits away from the UK. The information is based on the census and is self-reported, reflecting data as of 2011. Researchers interested in longitudinal analysis can utilise the year of arrival information to capture changes in length of residence over time.
mainlangprf11_cen, psspuk111_cen, natidbpuk11_cen	Note that these variables are based on the census and are self-reported at a single point in time. They may change over time and could be subject to measurement error, particularly as the data becomes more distant from 2011.

---

## How we dealt with data limitations

### *Preparing the data for analysis*

As outlined in Forth *et al.* (2022), the linking of the ASHE and the 2011 Census for England and Wales was conducted in two phases, referred to as 'Phase 1' and 'Phase 2'. The datasets produced from each phase are known as 'Drop 1' and 'Drop 2', respectively.

This project utilised the linked records from the second phase of the linkage process, i.e. Drop 2 of the ASHE – 2011 Census linked dataset.

Currently, Drop 1 of the ASHE – 2011 Census linked dataset, which contains the linked cases from the first phase, is available within the ONS SRS. However, Drop 2, which includes the full set of records from both phases (resulting in a higher match rate), as well as the linkage variables<sup>4</sup>, requires further processing by researchers before it is able to be used for analysis. To prepare the dataset, we employed the code files provided by the Wage and Employment Dynamics team within the ONS SRS, applying it to Drop 1. This allowed us to prepare Drop 2 ready for analysis.

### *Variables common to both ASHE and the 2011 Census*

In addition to the key variables described in the table above, our analysis utilised information on other employee characteristics, including age and gender.

The gender variable is common to both ASHE and the census and is therefore available from both sources in the linked dataset (`sex_ashe` and `sex_cen`). For a tiny fraction of employees (less than 1%), the two datasets disagree on recorded gender. In such cases, we treated these observations as missing.

The linked dataset also includes information on age from both ASHE and the Census (`age_ashe` and `age_cen`). For a small number of employees (less than 6%), the age recorded in two datasets differ. Given the continuous nature of this variable, we accepted a range of [-1,1] years, and treated observations as missing where the difference in age reported by the two sources falls outside this range.

Analysis where focus is on these variables should recognise that discrepancies between the two datasets could indicate errors in either source. Researchers who wish to include observations in their analysis – especially where the two sources disagree on an item – may find it useful to explore how sensitive their results are to the choice of dataset from which these variables are derived.

---

<sup>4</sup> See Forth *et al.* (2023) for a description of linkage variables included in the ASHE – 2011 Census linked dataset.

### *Using linkage and weighting variables*

The ASHE – 2011 Census linked dataset includes variables documenting aspects of the linkage process, such as a match quality score for each record (`score_link`). The dataset contains only observations with a match score of at least 0.82, which is considered as a good-quality link (Forth *et al.*, 2022). For our baseline analysis, we included all these observations to maximise sample size. However, we explored the sensitivity of our findings by focusing on cases with the highest match score of 1.0. Researchers may wish to assess the sensitivity of their results to the inclusion of cases with varying match quality.

It is possible to add to the dataset a weighting variable (`acew11_wt`)<sup>5</sup>, which can then be used to ensure estimates based on ASHE job records linked to the 2011 Census are representative of all jobs held by employees residing in England and Wales. While our baseline analysis was unweighted, where relevant, we explored the sensitivity of our findings to the application of these weights. Researchers may similarly evaluate how the use of these weights affects their results.

### *Lack of information on reasons for migration*

Previous UK evidence suggests that immigrant labour market patterns can vary significantly depending on their reason for migration (see, for example, Clark and Lindley, 2009). The ASHE – 2011 Census linked dataset does not include specific information on the reasons for migration for non-UK-born employees. However, it does provide the year of arrival in the UK (`yrarr_year_cen`) and the year in which employees started working for their organisation (`empsta_ashe`). Using this information, we identified a subgroup of labour market entrants—those whose employment start year coincides with their year of arrival in the UK—as likely having arrived in the UK primarily for work reasons. It is important to acknowledge that this group, however, might also include individuals who arrived in the UK for other reasons but found employment within the same year.

Additionally, we conducted analysis for the remaining non-UK-born employees. This group would include not only those who arrived for reasons such as education or to join family members, but also some labour market entrants who may have changed employers during their time in the UK.<sup>6</sup>

---

<sup>5</sup> This variable can be added to the data by merging the Drop 2 of the ASHE – 2011 Census linked dataset with the dataset containing weights (both of which are generated after running the code files shared by the Wage and Employments Dynamics team within the ONS SRS) using the anonymised personal and job (i.e., record reference number) identifiers (`adr_id_ashe` and `serno_ashe`) as merging variables.

<sup>6</sup> See Kaya (2024a) for findings from these analyses.

---

## Suggested improvements recommended to data owners

- Preparing Drop 2 of the ASHE – 2011 Census linked dataset requires a certain level of familiarity with the specific statistical software used to read the Wage and Employment Dynamic (WED) team’s code files. It would be helpful to provide Drop 2 of the ASHE – 2011 Census linked dataset directly within the ONS SRS in a format ready for analysis. This would eliminate the need for running the WED team’s code files on Drop 1, avoiding duplication and potential errors, and ensuring greater consistency in approach among users of the dataset. This would also assist new users of this data and make it more accessible to a wider range of prospective users.
- The current linked data covers the period up to 2018. Updating the linked dataset to include recent editions of ASHE (ONS, 2024)—or alternatively, providing the latest ASHE editions within the ONS SRS project space for users to append—would keep the data up to date. This would enable more comprehensive long-term analyses and allow users to examine the impact of recent events (e.g. the end of free movement), as well as changes and trends in the labour market.
- It would be beneficial to provide variable value labels and links to relevant lookups in the data documentation. This information would help users better understand the coding and classification of variables, reducing the risk of misinterpretation and enhancing the overall quality of their analyses.

## Additional data which would help to further develop the research

- At the time of writing, there were no equivalent linked datasets for other census years, although a link between ASHE and the 2021 Census for England and Wales is anticipated (Forth *et al.*, 2022). Extending the ASHE – 2011 Census to include a linkage with the ASHE – 2021 Census would be particularly valuable for future research. This extension would allow for an in-depth analysis of the labour market dynamics across various cohorts of immigrants over an extended period. It would also enable the examination of labour market performance of those who arrived in the UK and entered the labour market after 2011. Such linkage would also facilitate investigating the impact of changes in potentially time-varying factors such as English language proficiency, UK passport holding, and British national identity.
- To provide a comprehensive picture for the UK, obtaining comparable linked data for Scotland and Northern Ireland would be beneficial. The censuses for Scotland and Northern Ireland are governed by different arrangements compared to those for England and Wales. The Northern Ireland Statistics and Research Agency (NISRA) has successfully linked ASHE data to the 2011 Census for Northern Ireland, and this data is made available by NISRA. However, there are differences in the linkage approach compared to the ASHE

- 2011 Census for England and Wales.<sup>7</sup> At the time of writing, there has not yet been a planned linkage of the ASHE records to the 2011 Census for Scotland.
- To gain further insight into the variation in labour market experiences of different immigrant groups, triangulation with other datasets could be highly valuable. One area for deeper investigation is examining how labour market outcomes of immigrants vary by reasons for migration, such as for work, study, family or humanitarian reasons.

---

<sup>7</sup> See Gillespie *et al.* (2023) for a discussion.

---

## References

Clark, K. and Lindley, J., (2009). Immigrant Assimilation Pre and Post Labour Market Entry: Evidence from the UK Labour Force Survey. *Journal of Population Economics*, 22(1), pp.175-198.

Forth, J., Bryson, A., Phan, V., Ritchie, F., Singleton, C., Stokes, L., and Whittard, D., (2022). ASHE – Census 2011 Data Linkage: User Guide for Drop 2 of the ASHE –Census 2011 Dataset. Available at: <https://www.wagedynamics.com/wp-content/uploads/2023/01/ASHE-CEW11-User-Guide-Version-2.1-Drop-2.pdf>.

Forth, J., Bryson, A., Phan, V., Ritchie, F., Singleton, C., Stokes, L., and Whittard, D., (2023). Annual Survey of Hours and Earnings linked to 2011 Census – England and Wales. ADR UK Data Explained. Available at: [https://www.adruk.org/fileadmin/uploads/adruk/Documents/Data\\_Explained/Data\\_Explained\\_A\\_SHE-2011\\_Census\\_dataset.pdf](https://www.adruk.org/fileadmin/uploads/adruk/Documents/Data_Explained/Data_Explained_A_SHE-2011_Census_dataset.pdf).

Gillespie, A., McVicar, D., Rowland, N., and Shuttleworth-Queen's, I., (2023). Using the Earnings and Employees Study 2011 to Estimate Catholic-Protestant Wage Differentials. ADR NI Data Explained. Available at: [https://www.adruk.org/fileadmin/uploads/adruk/Documents/Data\\_Explained/Data-Explained-EES.pdf](https://www.adruk.org/fileadmin/uploads/adruk/Documents/Data_Explained/Data-Explained-EES.pdf).

Kaya, E. (2024a). Labour Market Performance of Immigrants: New Evidence from Linked Administrative Data. GLO Discussion Paper, No. 1418, Eses: Global Labor Organization. Available at: <https://ideas.repec.org/p/zbw/glodps/1418.html>.

Kaya, E. (2024b). Labour Market Differences between Immigrant and UK-born Employees: What is the Role of Employers? ADR UK Data Insight. Available at: <https://www.adruk.org/news-publications/publications-reports/labour-market-differences-between-immigrant-and-uk-born-employees-what-is-the-role-of-employers/>.

Kierans, D. (2021). Integration in the UK: Understanding the Data. Migration Observatory Report. University of Oxford.

Office for National Statistics [ONS]. (2022) International Migration, England and Wales: Census 2021. Statistical Bulletin. Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/bulletins/internationalmigrationenglandandwales/census2021>.

Office for National Statistics [ONS]. (2023). Released 03 July 2023, ONS SRS Metadata Catalogue, dataset, Annual Survey of Hours and Earnings linked to 2011 Census – England and Wales. DOI: <https://doi.org/10.57906/80f7-te97>.

Office for National Statistics [ONS]. (2024). Released 08 February 2024, ONS SRS Metadata Catalogue, dataset, Annual Survey of Hours and Earnings – GB. DOI: <https://doi.org/10.57906/x25d-4j96>.

## Disclaimer

The administrative data discussed in this Data Explained was made securely available through the ADR UK-funded Wage and Employment Dynamics programme. The data used in this research project comes from the Office for National Statistics (ONS) and was accessed through the ONS Secure Research Service (SRS).

The use of the ONS statistical data in this work does not imply the endorsement of the ONS or other data owners in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

## Acknowledgements

This work is supported by ADR UK (Administrative Data Research UK). ADR UK is a partnership transforming the way researchers access the UK's wealth of public sector data, to enable better informed policy decisions that improve people's lives. ADR UK is an Economic and Social Research Council (ESRC) investment (part of UK Research and Innovation). [Grant number: ES/Y001001/1]

## Contact

Name: Dr Ezgi Kaya

Email: [KayaE@cardiff.ac.uk](mailto:KayaE@cardiff.ac.uk)

